2-12-02 #3

LAW OFFICES
SUGHRUE, MION, ZINN, MACPEAK & SEAS, PLLC
2100 PENNSYLVANIA AVENUE, N.W.
WASHINGTON, DC 20037-3213
TELEPHONE (202) 293-7060
FACSIMILE (202) 293-7860
www.sughrue.com



J. Frank Osha, Esq.
Direct Dial (202) 663-7915
Email: fosha@sughrue.com

March 14, 2001

BOX PATENT APPLICATION
Commissioner for Patents
Washington, D.C. 20231

Re: Application of Kenji YAMANISHI and Hang LI
QUESTIONNAIRE ANALYSIS SYSTEM
Our Ref. Q63084

Dear Sir:

Attached hereto is the application identified above including 30 sheets of the specification, including the claims and abstract, 10 sheets of drawings, the executed Assignment and PTO 1595 form, and the executed Declaration and Power of Attorney. Also enclosed is an Information Disclosure Statement with form PTO-1449 and references.

The Government filing fee is calculated as follows:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Total claims | 13 | - 20 | = | | x | $18.00 | = | |
| Independent claims | 4 | - 3 | = | 1 | x | $80.00 | = | $80.00 |
| Base Fee | | | | | | | | $710.00 |
| **TOTAL FILING FEE** | | | | | | | | **$790.00** |
| Recordation of Assignment | | | | | | | | $40.00 |
| **TOTAL FEE** | | | | | | | | **$830.00** |

Checks for the statutory filing fee of $790.00 and Assignment recordation fee of $40.00 are attached. You are also directed and authorized to charge or credit any difference or overpayment to Deposit Account No. 19-4880. The Commissioner is hereby authorized to charge any fees under 37 C.F.R. §§ 1.16 and 1.17 and any petitions for extension of time under 37 C.F.R. § 1.136 which may be required during the entire pendency of the application to Deposit Account No. 19-4880. A duplicate copy of this transmittal letter is attached.

Priority is claimed from March 15, 2000 based on Japanese Application No. 2000-071657. The priority document is enclosed herewith.

Respectfully submitted,
SUGHRUE, MION, ZINN,
MACPEAK & SEAS, PLLC
Attorneys for Applicant

By: _____
J. Frank Osha
Registration No. 24,625

Yamanishi et al
Filed 3/14/01
Q63084
1 of 1

# 日 本 国 特 許 庁
PATENT OFFICE
JAPANESE GOVERNMENT



別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日
Date of Application: ２０００年 ３月１５日

出 願 番 号
Application Number: 特願２０００－０７１６５７

出 願 人
Applicant (s): 日本電気株式会社

２０００年１２月 １日

特 許 庁 長 官
Commissioner,
Patent Office 及 川 耕 造

出証番号 出証特２０００－３１００２１６

【書類名】　特許願
【整理番号】　33509710
【あて先】　特許庁長官殿
【国際特許分類】　G06F 17/00
【発明者】
　【住所又は居所】　東京都港区芝五丁目７番１号　日本電気株式会社内
　【氏名】　山西　健司
【発明者】
　【住所又は居所】　東京都港区芝五丁目７番１号　日本電気株式会社内
　【氏名】　李　航
【特許出願人】
　【識別番号】　000004237
　【氏名又は名称】　日本電気株式会社
【代理人】
　【識別番号】　100088890
　【弁理士】
　【氏名又は名称】　河原　純一
【手数料の表示】
　【予納台帳番号】　009690
　【納付金額】　21,000円
【提出物件の目録】
　【物件名】　明細書　1
　【物件名】　図面　1
　【物件名】　要約書　1
　【包括委任状番号】　9001717
【プルーフの要否】　要

【書類名】　　明細書

【発明の名称】アンケート回答分析システム

【特許請求の範囲】

【請求項１】自然言語による自由回答記述を含むアンケート回答文を入力する手段と、

前記アンケート回答文を送信するネットワークと、

前記送信されたアンケート回答文を蓄積するデータベースと、

該データベースから前記アンケート回答文を読み出して、前記アンケート回答文を分類するルールを学習するテキスト分類エンジンと

を備えることを特徴とするアンケート回答分析システム。

【請求項２】自然言語による自由回答記述を含むアンケート回答文を入力する手段と、

前記アンケート回答文を蓄積するデータベースと、

該データベースから前記アンケート回答文を読み出して、前記アンケート回答文を分類するルールを学習するテキスト分類エンジンと

を備えることを特徴とするアンケート回答分析システム。

【請求項３】自然言語による自由回答記述を含むアンケート回答文を入力する手段と、

前記アンケート回答文を送信するネットワークと、

前記送信されたアンケート回答文を蓄積するデータベースと、

該データベースから前記アンケート回答文を読み出して、前記アンケート回答文を分類するルールを学習するテキスト分類エンジンと、

該ルールを要求者からの要求に応じて前記ネットワークを通じて配信する手段と

を備えることを特徴とするアンケート回答分析システム。

【請求項４】前記テキスト分類エンジンが、前記データベースに蓄積されたアンケート回答文の全文に対して形態素解析を行う形態素解析手段と、カテゴリとテキストとを指定させるカテゴリ・テキスト指定手段と、前記データベースから読み込んだ複数のアンケート回答文に対して属性選択を行う属性選択手段と、前記

属性選択手段により属性選択された単語を基にテキストとカテゴリとの対応を表すルールを学習するルール学習手段と、前記ルール学習手段により学習されたルールを出力するルール出力手段とを含む請求項１，請求項２，または請求項３記載のアンケート回答分析システム。

【請求項５】前記属性選択手段が、テキスト中に出現する単語の各々について、該単語の出現を考慮しない場合のテキスト集合の確率的コンプレキシティと考慮する場合のテキスト集合の確率的コンプレキシティとの差ΔＳＣ（ω）を計算する手順と、該差ΔＳＣ（ω）がしきい値τより大きければ属性として選択する手順とを実行する請求項４記載のアンケート回答分析システム。

【請求項６】前記ルール学習手段が、テキストの集合を、（$d_1$，$c_1$），（$d_2$，$c_2$），…，（$d_m$，$c_m$）［ただし、各$d_i$は多値の離散ベクトル$d_i=(\omega_{i1}, \omega_{i2}, \cdots, \omega_{in})$（ｉ＝１，…，ｍ）、$\omega_{ij}$は属性選択で得られた単語$\omega_j$（ｊ＝１，…，ｎ）がｉ番目のテキストに現れれば１、そうでなければ０の値をとる、$c_i$はｉ番日のテキストに対応するカテゴリの値（ラベル）を表し、各$c_i$は所与のカテゴリに属するならば１、そうでなければ０の値をとる、ｍはテキストの数］の表現に置き換えて成形する手順と、拡張型確率的コンプレキシティ最小原理または確率的コンプレキシティ最小化原理などの情報量規準を用いてｉｆ－ｔｈｅｎ－ｅｌｓｅ型のルールを選択して確率的決定リストに順番に追加していく成長処理を行う手順と、前記確率的決定リストの最後からルールを１つずつ取り除き拡張型確率的コンプレキシティ最小原理の観点からみてこれを取り除かないほうがよいところまで続ける刈り込み処理を行う手順とを実行する請求項４記載のアンケート回答分析システム。

【請求項７】コンピュータを、前記データベースに蓄積されたアンケート回答文の全文に対して形態素解析を行う形態素解析手段，前記テキスト分類エンジンに対してカテゴリとテキストとを指定させるカテゴリ・テキスト指定手段，前記データベースから読み込んだ複数のアンケート回答文に対して属性選択を行う属性選択手段，前記属性選択手段により属性選択された単語を基にテキストとカテゴリとの対応を表すルールを学習するルール学習手段，および前記ルール学習手段により学習されたルールを出力するルール出力手段として機能させるためのプロ

グラムを記録した記録媒体。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】

本発明はアンケート回答分析システムに関し、特にテキスト自動分類技術，自然言語処理技術，およびネットワーク利用技術を使用するアンケート回答分析システムに関する。

【０００２】

【従来の技術】

インターネット等のネットワークを通じて得られる、自然言語による自由回答記述を含むアンケート回答文から全体的な特徴や傾向を導き出す作業は、従来の多くは人手で行われてきた。 株式会社電通が開発したＤＥ－ＦＡＣＴＯ(パンフレット)や富士通株式会社のＫｅｙｗｏｒｄ Ａｓｓｏｃｉａｔｏｒ(渡部勇：発散的思考支援システム「Ｋｅｙｗｏｒｄ Ａｓｓｏｃｉａｔｏｒ」第二版，計測自動制御学会第１５回システム工学部会研究会資料，１９９４年７月)，ＨＩＰＳ等のテキストマイニングツール(渡辺，三末，新田，杉山： ハイブリッド発想支援システム「ＨＩＰＳ」，計測自動制御学会第１５回システム工学部会研究会資料，１９９５年１月)は、テキスト情報から言葉の関係性を導き出すのに用いられた。しかし、これらのツールは、ルールという形式でアンケート回答文の特徴を表現するものではなかった。

【０００３】

また、自然言語による自由回答記述を含むアンケート回答文をインターネット等のネットワークを通じて自動的に収集して分析し、場合によっては要求者に分析結果を配信するようなシステムやサービスはこれまで存在しなかった。たとえば、特開平１１－０６６０３６号公報，特開平１１－１４３８５６号公報等には、ネットワークを通してアンケートを実施し、回答をデータベースに蓄積する技術が開示されているが、ルールという形式でアンケート回答文の特徴を抽出することまでは行われていない。

【０００４】

【発明が解決しようとする課題】

上述した従来の人手によるアンケート回答分析では、アンケート回答文の数が大量になると、非効率で手に負えないという問題点があった。

【０００５】

また、ＤＥ－ＦＡＣＴＯやＨＩＰＳ等のテキストマイニングツールでは、ルールという形式でアンケート回答文の特徴を抽出できないので、コンパクトで分かりやすい知識の提示という点で十分ではないという問題点があった。

【０００６】

さらに、検索等で用いられているテキスト分類ツールは、アンケート回答分析にも有力なツールであるが、自然言語による自由回答記述を含むアンケート回答文の分析に用いられている前例はなかった。

【０００７】

本発明の第１の目的は、テキスト分類エンジンを用いることによって、自然言語による自由回答記述を含むアンケート回答文からコンパクトで分かりやすいルール形式の知識を自動的に提示することを可能にするアンケート回答分析システムを提供することにある。

【０００８】

また、本発明の第２の目的は、上記に加えて、ネットワークを通じて集めたアンケート回答文からルール形式の知識を自動的に抽出し、要求者に分析結果を配信するアンケート回答分析システムを提供することにある。

【０００９】

【課題を解決するための手段】

本発明のアンケート回答分析システムは、自然言語による自由回答記述を含むアンケート回答文を入力する手段と、前記アンケート回答文を送信するネットワークと、前記送信されたアンケート回答文を蓄積するデータベースと、該データベースから前記アンケート回答文を読み出して、前記アンケート回答文を分類するルールを学習するテキスト分類エンジンとを備えることを特徴とする。

【００１０】

また、本発明のアンケート回答分析システムは、自然言語による自由回答記述を

含むアンケート回答文を入力する手段と、前記アンケート回答文を蓄積するデータベースと、該データベースから前記アンケート回答文を読み出して、前記アンケート回答文を分類するルールを学習するテキスト分類エンジンとを備えることを特徴とする。

【００１１】

さらに、本発明のアンケート回答分析システムは、自然言語による自由回答記述を含むアンケート回答文を入力する手段と、 前記アンケート回答文を送信するネットワークと、前記送信されたアンケート回答文を蓄積するデータベースと、該データベースから前記アンケート回答文を読み出して、前記アンケート回答文を分類するルールを学習するテキスト分類エンジンと、該ルールを要求者からの要求に応じて前記ネットワークを通じて配信する手段とを備えることを特徴とする。

【００１２】

【発明の実施の形態】

以下、本発明の実施の形態について図面を参照して詳細に説明する。

【００１３】

（１）第１の実施の形態

図１は、本発明の第１の実施の形態に係るアンケート回答分析システムの全体構成を示すブロック図である。本実施の形態に係るアンケート回答分析システムは、回答者コンピュータ１１１～１１Ｎ（Ｎは正整数）と、ネットワーク１２と、データベース１３と、テキスト分類エンジン１４とから、その主要部が構成されている。

【００１４】

回答者コンピュータ１１１～１１Ｎは、コンピュータ，携帯情報端末，携帯電話機等の各種のメッセージ，メール等の送信機能を持つ機器が該当し、ネットワーク１２に接続されている。

【００１５】

ネットワーク１２には、有線，無線を問わない各種の公衆回線網，専用線網や、ＬＡＮ（Ｌｏｃａｌ　Ａｒｅａ　Ｎｅｔｗｏｒｋ）等の各種のネットワークが含

まれる。

【００１６】

データベース１３は、ネットワーク１２に接続されており、回答者コンピュータ１１１～１１Ｎからネットワーク１２を通じて送信された、複数のアンケート回答者からのアンケート回答文を蓄積する。

【００１７】

テキスト分類エンジン１４は、データベース１３から複数のアンケート回答文を読み出して、アンケート回答文を分類するルールを導き出し、ルールを要求者に出力する。テキスト分類エンジン１４は、データベース１３に蓄積されたアンケート回答文の全文に対して形態素解析を行う形態素解析手段１５と、テキスト分類エンジン１４に対してカテゴリとテキストとを指定させるカテゴリ・テキスト指定手段１６と、データベース１３から読み込んだ複数のアンケート回答文に対して属性選択を行う属性選択手段１７と、属性選択手段１７により属性選択された単語を基にテキストとカテゴリとの対応を表すルールを学習するルール学習手段１８と、ルールを出力するルール出力手段１９とを含んで構成されている。

【００１８】

なお、テキスト分類エンジン１４は、カテゴリとテキストとの間の対応関係を分類則として学習するエンジンであり、たとえば、Ｌｉ ａｎｄ Ｙａｍａｎｉｓｈｉにより提案されているエンジン（ｈ．Ｌｉ ａｎｄ ｋ．Ｙａｍａｎｉｓｈｉ：Ｔｅｘｔ Ｃｌａｓｓｉｆｉｃａｔｉｏｎ Ｕｓｉｎｇ ＥＳＣ－ｂａｓｅｄ Ｓｔｏｃｈａｓｔｉｃ Ｄｅｃｉｓｉｏｎ Ｌｉｓｔｓ，Ｐｒｏｃｃｅｄｉｎｇｓ ｏｆ １９９９ Ｉｎｔｅｒｎａｔｉｏｎａｌ Ｃｏｎｆｅｒｅｎｃｅ ｏｎ Ｉｎｆｏｒｍａｔｉｏｎ ＆ Ｋｎｏｗｌｅｄｇｅ Ｍａｎａｇｅｍｅｎｔ，ｐｐ：１２２－１３０，１９９９）を用いることができる。このテキスト分類エンジン１４は、基本的に特許第２５８１１９６号に開示された「決定リストの生成方法及び装置」の方式を利用したものである。

【００１９】

図２は、データベース１３に格納されるアンケート回答文の集合の構造を示す。各列はアンケート項目を表し、各行は１人分のアンケート回答文を表す。

【００２０】

図３を参照すると、テキスト分類エンジン１４の処理は、形態素解析ステップ３１と、テキストとカテゴリとの指定ステップ３２と、属性選択ステップ３３と、ルール学習ステップ３４と、ルール出力ステップ３５とからなる。

【００２１】

図４を参照すると、属性選択ステップ３３のより詳しい処理は、ΔＳＣ（ω）計算ステップ４１と、属性選択ステップ４２とからなる。

【００２２】

図５を参照すると、ルール学習ステップ３４のより詳しい処理は、データ成形ステップ５１と、成長処理ステップ５２と、刈り込み処理ステップ５３とからなる。

【００２３】

図６は、テキスト分類エンジン１４による分析結果であるルール形式の知識（確率的決定リスト）の一例を示す図である。

【００２４】

図７は、テキスト分類エンジン１４による分析結果であるルール形式の知識（確率的決定リスト）の他の例を示す図である。

【００２５】

次に、このように構成された第１の実施の形態に係るアンケート回答分析システムの動作について説明する。

【００２６】

アンケート回答者が、回答者コンピュータ１１１～１１Ｎからアンケート回答文を送信すると、アンケート回答文はネットワーク１２を通じてデータベース１３に蓄積される。ここで、アンケート回答者の数は、Ｎ人とする。この際、アンケート回答文は、自然言語による自由回答記述を含むものであってよい。

【００２７】

テキスト分類エンジン１４は、まず、形態素解析手段１５により、データベース１３に蓄積されたアンケート回答文の全文に対して形態素解析を行う（ステップ３１）。

【００２８】

次に、テキスト分類エンジン１４は、カテゴリ・テキスト指定手段１６により、オペレータにアンケート回答文中のカテゴリとテキストとを指定させる（ステップ３２）。ここで、カテゴリの指定とは、１つの列の回答に注目して分類するものである。例えば、図２の１列目に注目して、その回答を「Ａ社」と「Ａ社以外」とに分類することが、カテゴリ指定である。また、テキスト指定とは、カテゴリ指定に用いられた列を除いて、分析に用いる列を指定することである。 例えば、図２の２列目を選択してテキスト指定を行う。

【００２９】

続いて、テキスト分類エンジン１４は、属性選択手段１７により、データベース１３から読み込んだ複数のアンケート回答文に対して属性選択を行う（ステップ３３）。ここで、属性選択とは、テキストとカテゴリとの対応を表すのに重要な単語を選択することである。

【００３０】

次に、テキスト分類エンジン１４は、ルール学習手段１８により、属性選択された単語を基にテキストとカテゴリとの対応を表すルールを学習する（ステップ３４）。例えば、上記のカテゴリ指定とテキスト指定とをしたときに、図６のようなルールが得られる。

【００３１】

図６のルールは、先ず最初の行を読んで、テキストの中で「使いやすい」という単語が現れていれば、そのような回答をした人の９２．０％がハイテク企業としてＡ社を想起している、ということを示す。もし、「使いやすい」という単語が現れていなければ、次に、「未来」と「プライベート」という単語が同時に現れているかどうかをチェックし、現れていれば、そのような回答をした人の８７．２％がハイテク企業としてＡ社を想起している、ということを示す。以下同様に、ｉｆ－ｔｈｅｎ－ｅｌｓｅ型のルールに従って、上から下へと条件文を読んで行く。このようなルールは、ハイテク企業とハイテク感との間の関係を分かりやすくコンパクトに示すものである。

【００３２】

また、別の会社であるＢ社をとりあげて、「Ｂ社」と「Ｂ社以外」とにカテゴリ指定したときに、同様な手順で図７のルールが得られたとする。

【００３３】

図７のＢ社のルールを図６のＡ社のルールと比較すると、Ａ社をハイテク企業と想起する人のハイテク感は、使い勝手や嗜好品感覚を重視するのに対し、Ｂ社をハイテク企業と想起する人のハイテク感は効率性を重視していることがわかる。このように、ルールの比較により、アンケート回答分析が容易になる。

【００３４】

最後に、テキスト分類エンジン１４は、ルール出力手段１９により、分析結果のルール形式の知識を要求者に出力する（ステップ３５）。

【００３５】

ここで、ルール形式の知識の例として確率的決定リストをとり上げ、これを生成するための属性選択ステップ３３およびルール学習ステップ３４についてより詳しく説明する。

【００３６】

確率的決定リストとは、図６に示すようなｉｆ－ｔｈｅｎ型の確率的ルールの順序付きリストのことである。各確率的ルールは、”ｃ＝１←ｔ（確率ｐ）”の形をとる。ここで、ｃ＝１は分類の決定、ｔは分類の条件、（確率ｐ）は確率を表す。

【００３７】

まず、属性選択ステップ３３について詳しく述べる。

【００３８】

属性選択とは、与えられたカテゴリ（例えば、’Ａ社’と’Ａ社でない’）に対して、そのカテゴリと深く関係する単語を集めることである。具体的には、図４に示すように、ステップ４１で、テキストに出現する単語ωの各々について、その単語ωの出現を考慮しない場合のテキスト集合の確率的コンプレキシティ（ＳＣとかく）と考慮する場合のＳＣとの差ΔＳＣ（ω）を計算し、ステップ４２で、その差ΔＳＣ（ω）が、与えられたしきい値τより大きければ単語ωを属性として選択する。

【００３９】

以下、ＳＣの具体的な計算の仕方を述べる。入力されたアンケート回答文中のテキストの集合を

（$d_1$，$c_1$），（$d_2$，$c_2$），…，（$d_m$，$c_m$）

と表す。ここで、$d_i$はｉ番目のテキストを表し、ｉ番目のテキストに現れる単語の列として表現される。$c_i$はｉ番目のテキストに対応するカテゴリの値（ラベル）を表し、各$c_i$は所与のカテゴリ（’Ａ社’）に属するならば１、そうでなければ（’Ａ社でない’）０の値をとる。ｍはテキストの数である。

【００４０】

さらに、ラベル列を$c^m = c_1, \cdots, c_m$、テキスト列を$d^m = d_1, \cdots, d_m$と書く。ラベル列$c^m$のＳＣを、数１のように計算する。

【００４１】

【数１】

$$SC(c^m) = mH\left(\frac{m^+}{m}\right) + \frac{1}{2}\log\frac{m}{2\pi} + \log\pi$$

【００４２】

ここでは、$m^+$はラベル列$c^m$中で値が１であるラベルの数を表す。また、ｌｏｇは自然対数を表し、数２とする。

【００４３】

【数２】

$$H(z) \stackrel{def}{=} -z\log z - (1-z)\log(1-z)$$

【００４４】

たとえば、Ｊ．Ｒｉｓｓａｎｅｎ，Ｆｉｓｈｅｒ ｉｎｆｏｒｍａｔｉｏｎ ａｎｄ ｓｔｏｃｈａｓｔｉｃ ｃｏｍｐｌｅｘｉｔｙ．ＩＥＥＥ Ｔｒａｎｓ．ｏｎ Ｉｎｆｏｍａｔｉｏｎ Ｔｈｅｏｒｙ，４２（１）：４０－４７（１９９６）に述べられているように、ＳＣ（$c^m$）は、ラベル列$c^m$を与えられたモ

デル（ここではベルヌイモデル）を用いて記述するのに必要な最短の記述長という意味をもつ。また、

【外１】

$c^{m_\omega}$

は、対応するテキストｄ$_i$の中に単語ωが現れるラベルｃ$_i$からなるラベル列であるとする。ここに、ｍ$_\omega$は、

【外２】

$c^{m_\omega}$

におけるラベルの数であるとする。すると、

【外３】

$c^{m_\omega}$

のＳＣの値を、数３のように計算することができる。

【００４５】

【数３】

$$SC(c^{m_\omega}) = m_\omega H\left(\frac{m_\omega^{+}}{m_\omega}\right) + \frac{1}{2}\log\frac{m_\omega}{2\pi} + \log\pi$$

【００４６】

ここでは、ｍ$_\omega^{+}$は、

【外４】

$c^{m_\omega}$

における値が１であるラベルの数を表す。一方、

【外５】

$c^{m_{\neg\omega}}$

は、対応するテキストｄ$_i$に単語ωが現れないラベルｃ$_i$からなるラベル列であるとする。ｍ$_\omega$は、

【外６】

$c^{m_{\neg\omega}}$

におけるラベルの数である。

【００４７】

【外７】

$c^{m_{\neg\omega}}$

のＳＣの値を、数４のように計算することができる。

【００４８】

【数４】

$$SC(c^{m_{\neg\omega}}) = m_{\neg\omega} H\left(\frac{m_{\neg\omega}^{+}}{m_{\neg\omega}}\right) + \frac{1}{2}\log\frac{m_{\neg\omega}}{2\pi} + \log\pi$$

【００４９】

単語ωの出現を考慮しない場合のＳＣと考慮する場合のＳＣとの差ΔＳＣ（ω）は、数５のように計算される。

【００５０】

【数５】

$$\Delta SC(\omega) = \frac{1}{m}\left(SC(c^{m}) - (SC(c^{m_\omega}) + SC(c^{m_{\neg\omega}}))\right)$$

$$= \left[H\left(\frac{m^{+}}{m}\right) - \frac{m_\omega}{m} H\left(\frac{m_\omega^{+}}{m_\omega}\right) - \frac{m_{\neg\omega}}{m} H\left(\frac{m_{\neg\omega}^{+}}{m_{\neg\omega}}\right)\right]$$

$$-\left[\frac{1}{2m}\log\frac{m_\omega m_{\neg\omega}\pi}{2m}\right]$$

【００５１】

差ΔＳＣ（ω）の大きい単語ωは、与えられたカテゴリによく現れる、あるいはほとんど現れない単語である。それらの単語がそのカテゴリと深く関係するとみることができる。そこで、τを与えられた閾値とし、ΔＳＣ（ω）＞τなる単語ωを属性として選択する。

【００５２】

次に、ルール学習ステップ３４について詳しく述べる。

【００５３】

いま、属性選択された単語がｎ個あるとし、$\omega_1$，…，$\omega_n$とする。ステップ５１で、まず、入力されたテキストの集合を、以下の表現に置き換える。

（$d_1$，$c_1$），（$d_2$，$c_2$），…，（$d_m$，$c_m$）

【００５４】

ここで、各$d_i$は、２値ベクトル（一般的に多値の離散ベクトル）

$d_i=(\omega_{i1}, \omega_{i2}, \cdots, \omega_{in})$（ｉ＝１，…，ｍ）

を表す。ここで、$\omega_{ij}$は、属性選択で得られた単語$\omega_j$がｉ番目のテキストに現れれば１、そうでなければ０の値をとる（ｊ＝１，…，ｎ）。$c_i$はｉ番目のテキストに対応するカテゴリの値（ラベル）を表し、各$c_i$は所与のカテゴリに属するならば１、そうでなければ０の値をとる。ｍはテキストの数である。

【００５５】

ステップ５２では、ｉｆ－ｔｈｅｎ－ｅｌｓｅ型のルールを選択し、確率的決定リストＡに順番に追加していく。これを「成長」とよぶ。ルールの選択には、例えば、拡張型確率的コンプレキシティ（ＥＳＣとかく）最小原理を用いる。

【００５６】

これは、以下のように行う。ｋを与えられた正の整数とする。属性選択された単語ωを基に可能なすべてのｋ項（単語ωのｋ個までの連言）の集合をＴとする。次に、集合Ｔの項ｔの中からテキストに一度も現れないものを取り除く。また、空の確率的決定リストＡを用意する。次に、ＥＳＣの値の減少分がもっとも大きいルールを確率的決定リストＡに順次追加する。

【００５７】

ここで、ＥＳＣの計算は、以下のように行う。全データ集合Ｄを

（$d_1$，$c_1$），（$d_2$，$c_2$），…，（$d_m$，$c_m$）

の形のデータの集合とし、ラベル列を$c^m = c_1, \cdots, c_m$とする。ラベル列$c^m$のＥＳＣの値は、数６のように近似計算することができる。

【００５８】

【数６】

$$ESC(c^m) = Loss(c^m) + \lambda\sqrt{m \log m}$$

【００５９】

これは、Ｋ．Ｙａｍａｎｉｓｈｉ，Ａ ｄｅｃｉｓｉｏｎ－ｔｈｅｏｒｅｔｉｃ ｅｘｔｅｎｓｉｏｎ ｏｆ ｓｔｏｃｈａｓｔｉｃ ｃｏｍｐｌｅｘｉｔｙ，ａｎｄ ｉｔｓ ａｐｐｌｉｃａｔｏｎｓ ｔｏ ｌｅａｒｎｉｎｇ，ＩＥＥＥ Ｔｒａｎｓ．Ｉｎｆｏｒｍ．Ｔｈｅｏｒｙ，４４，１４２４－１４３９（１９９８）の論文で提案されたオリジナルのＥＳＣの１つの近似形である。

【００６０】

ここでは、λは正の定数を表し、Ｌｏｓｓ（$c^m$）はデフォルトの分類を行う際の誤りの数を表す。デフォルトの分類とは、たとえば、すべてのラベルが０であると仮定することである。ｔは集合Ｔの中の項であるとする。

【００６１】

【外８】

$c^{m_t}$

は、対応するテキスト$d_i$において項ｔが真になるラベル$c_i$からなるラベル列であるとする。ここに、$m_t$は、

【外９】

$c^{m_t}$

におけるラベルの数であるとする。

【００６２】

【外１０】

$$Loss\ (c^{m_t})$$

は、項ｔによる分類を行う際の誤りの数であるとする。一方、

【外１１】

$$c^{m_{\neg\omega}}$$

は、対応するテキストｄ$_i$において項ｔが偽になるラベルｃ$_i$からなるラベル列であるとする。ここに、$m_{\neg t}$は、

【外１２】

$$c^{m_{\neg\omega}}$$

におけるラベルの数であるとする。¬ｔは項ｔの否定を表す。

【００６３】

【外１３】

$$Loss\ (c^{m_{\neg t}})$$

は、¬ｔによる分類を行う際の誤りの数であるとする。

【００６４】

【外１４】

$$c^{m_t}$$

と

【外１５】

$$c^{m_{\neg t}}$$

とのＥＳＣの値を、それぞれ数７および数８のように計算することができる。

【００６５】

【数７】

$$ESC(c^{m_t}) = Loss(c^{m_t}) + \lambda\sqrt{m_t \log m_t}$$

【００６６】

【数８】

$$ESC(c^{m_{\neg t}}) = Loss(c^{m_{\neg t}}) + \lambda\sqrt{m_{\neg t} \log m_{\neg t}}$$

【００６７】

項ｔによる分類を行う場合、ＥＳＣの値の減少分ΔＥＳＣ（ｔ）を、数９で計算する。

【００６８】

【数９】

$$\begin{aligned}\Delta ESC(t) &= ESC(c^m) - (ESC(c^{m_t}) + ESC(c^{m_{\neg t}})) \\ &= \left[Loss(c^m) - Loos(c^{m_t}) - Loos(c^{m_{\neg t}})\right] \\ &\quad + \left[\lambda(\sqrt{m \log m} - \sqrt{m_t \log m_t} - \sqrt{m_{\neg t} \log m_{\neg t}})\right]\end{aligned}$$

【００６９】

ＥＳＣ最小原理では、ΔＥＳＣ（ｔ）が最小になるように項ｔを選択する。そのようなｔ＝ｔ＊が選ばれたならば、これを真とするような全データ集合Ｄのデータの数を

【外１６】

$m_{t^*}$

とし、そのようなデータのうち、かりに数が多い方のラベルをｃ＝１として、ｃ＝１であった数を

【外１７】

$m_{t^*}^{+}$

と、ｃ＝０であった数を

【外１８】

$$m_{t^*}^{-}$$

とする。そこで、ルール”ｃ＝１←ｔ＊（確率ｐ）”を確率的決定リストＡに追加する。ここで、確率値ｐはラプラス推定の方法を用いて、例えば、数１０により計算する。

【００７０】

【数１０】

$$(m_{t^*}^{+} + 0.5)/(m_{t^*} + 1)$$

【００７１】

集合Ｔから項ｔ＊を除いたものを新しい集合Ｔとし、全データ集合Ｄから項ｔ＊を真にする全てのデータを除いたものを新しい全データ集合Ｄとして、全データ集合Ｄが空になるまで、同じ操作を繰り返す。なお、上記に用いた基準ＥＳＣの代わりに、属性選択に用いた基準ＳＣを用いることもできる。

【００７２】

次に、ステップ５３では、ステップ５２で得られた確率的決定リストＡが学習データに過度に適合することがあるので、確率的決定リストＡの最後からルールを１つずつ取り除き、ＥＳＣ最小原理の観点からみてこれを取り除かないほうがよいところまで続ける。これを「刈り込み」とよぶ。

【００７３】

この場合のＥＳＣ最小原理の適用の仕方を、以下に述べる。まず、ラベル列ｃ$^{m}$の確率的決定リストＡに対するＥＳＣの値を、数１１のように、確率的決定リストＡにおけるすべての項ｔに対するＥＳＣの値の和として定義する。

【００７４】

【数１１】

$$ESC(c^m|A) = \sum_t ESC(c^{m_t})$$

【００７５】

ただし、

【外１９】

$ESC(c^{m_t})$

は、数７のように計算する。次に、ラベル列ｃ$^{m}$と確率的決定リストＡとの全ＥＳＣ値を、数１２のように定義する。

【００７６】

【数１２】

$$ESC(c^m : A) = ESC(c^m|A) + \lambda' L(A)$$
$$= \sum_t Loss(c^{m_t}) + \lambda \sum_t \sqrt{m_t \log m_t} + \lambda' L(A)$$

【００７７】

ここでは、λ’は正の定数を表す。Ｌ（Ａ）は、確率的決定リストＡを符号化する時の符号長である。具体的には、Ｌ（Ａ）＝ｌｏｇＴ＋ｌｏｇ（Ｔ－1）＋…＋ｌｏｇＴ（Ｔ－ｉ＋１）と計算する。ただし、Ｔは可能な項ｔの数で、ｉは確率的決定リストＡにおけるルールの数である。

【００７８】

Ａは刈り込み前の確率的決定リストを、Ａ’は刈り込み後の確率的決定リストを表すものとする。

ＥＳＣ（ｃ$^{m}$：Ａ）≦ＥＳＣ（ｃ$^{m}$：Ａ’）

【００７９】

すなわち、

ＥＳＣ（ｃ$^{m}$｜Ａ’）－ＥＳＣ（ｃ$^{m}$｜Ａ）≧λ’（Ｌ（Ａ）－Ｌ（Ａ’））

が成り立つ限り、刈り込み処理を続け、この条件が満足されなくなったか、刈り込むべきルールが無くなった時点で得られている確率的決定リストＡを出力する。ＥＳＣが全体として小さい確率的決定リストＡが出力されることになる。

【００８０】

第１の実施の形態に係るアンケート回答分析システムでは、ネットワーク１２を通じて収集した自然言語による自由回答記述を含むアンケート回答文から自動的に分析結果のルールを導くことができる。

【００８１】

第１の実施の形態に係るアンケート回答分析システムにおいて、テキスト分類エンジン１４として、例えば、Ｌｉ ａｎｄ Ｙａｍａｎｉｓｈｉにより提案されているエンジン（ｈ．Ｌｉ ａｎｄ ｋ．Ｙａｍａｎｉｓｈｉ：Ｔｅｘｔ Ｃｌａｓｓｉｆｉｃａｔｉｏｎ Ｕｓｉｎｇ ＥＳＣ－ｂａｓｅｄ Ｓｔｏｃｈａｓｔｉｃ Ｄｅｃｉｓｉｏｎ Ｌｉｓｔｓ，Ｐｒｏｃｅｅｄｉｎｇｓ ｏｆ １９９９ Ｉｎｔｅｒｎａｔｉｏｎａｌ Ｃｏｎｆｅｒｅｎｃｅ ｏｎ Ｉｎｆｏｒｍａｔｉｏｎ ＆ Ｋｎｏｗｌｅｄｇｅ Ｍａｎａｇｅｍｅｎｔ，ｐｐ：１２２－１３０，１９９９）を用いると、Ｏ（$n^k m$）の計算量で、高速にアンケート回答文からルールを導き出すことができる。ここに、ｎはアンケート回答文から属性選択された単語の数、ｍはアンケート回答文の数、ｋはルール１つの条件にとられる連言に含まれる単語の最大数である。よって、効率的な自動アンケート回答分析が可能となる。また、得られたルールは特定のカテゴリに属するアンケート回答文をｉｆ－ｔｈｅｎ－ｅｌｓｅ型のルール形式で、コンパクトにかつ分かりやすく表現している。

【００８２】

第１の実施の形態に係るアンケート回答分析システムは、例えば、以下のようなビジネスとして展開できる。企業に関するイメージ調査，特定の商品やサービス等のアンケートを請け負って、図２のようなアンケート項目でアンケートをネットワーク１２上で実施し、ネットワーク１２を通じてオンラインで集められた自然言語による自由回答記述を含むアンケート回答文をデータベース１３に蓄え、そこからアンケート回答文を呼び出して、テキスト分類エンジン１４を使用する

ことにより得られるルールを分析結果として要求者に販売する。

【００８３】

（２）第２の実施の形態

図８は、本発明の第２の実施の形態に係るアンケート回答分析システムの全体構成を示すブロック図である。本実施の形態に係るアンケート回答分析システムは、アンケート回答入力手段８１と、データベース８２と、テキスト分類エンジン８３とから、その主要部が構成されている。

【００８４】

アンケート回答入力手段８１は、ネットワークを介さずにデータベース８２と直接接続されている。

【００８５】

データベース８２は、複数のアンケート回答者からのアンケート回答文を蓄積する。

【００８６】

テキスト分類エンジン８３は、図１に示した第１の実施の形態に係るアンケート回答分析システムにおけるテキスト分類エンジン１４と全く同様のものである。したがって、対応する部分には同一符号を付して、それらの詳しい説明を省略する。

【００８７】

次に、このように構成された第２の実施の形態に係るアンケート回答分析システムの動作について説明する。

【００８８】

アンケート回答入力手段８１は、ネットワークを介さずにデータベース８２と直接接続して、自然言語による自由回答記述を含むアンケート回答文を入力する。

【００８９】

データベース８２は、複数のアンケート回答者からのアンケート回答文を蓄積する。

【００９０】

テキスト分類エンジン８３は、データベース８２から複数のアンケート回答文を読み出して、アンケート回答文を分類するルールを導き出し、分析結果のルールを要求者に出力する。なお、テキスト分類エンジン８３の動作の詳細は、第１の実施の形態に係るアンケート回答分析システムにおけるテキスト分類エンジン１４の場合と全く同様であるので、その詳しい説明を割愛する。

【００９１】

第２の実施の形態に係るアンケート回答分析システムは、例えば、以下のようなビジネスとして展開できる。企業に関するイメージ調査，特定の商品やサービス等のアンケートを請け負って、図２のようなアンケート項目でアンケートを実施し、自然言語による自由回答記述を含むアンケート回答文を一度に回収して、これをデータベース８２で蓄え、そこからアンケート回答文を呼び出して、テキスト分類エンジン８３を使用することにより得られる分析結果を要求者に販売する。

【００９２】

（３）第３の実施の形態

図９は、本発明の第３の実施の形態に係るアンケート回答分析システムの全体構成を示すブロック図である。本実施の形態に係るアンケート回答分析システムは、回答者コンピュータ９１１～９１Ｎと、ネットワーク９２と、データベース９３と、テキスト分類エンジン９４と、要求者コンピュータ９５とから、その主要部が構成されている。

【００９３】

回答者コンピュータ９１１～９１Ｎは、コンピュータ，携帯情報端末，携帯電話機等の各種のメッセージ，メール等の送信機能を持つ機器が該当し、ネットワーク９２に接続されている。

【００９４】

ネットワーク９２には、有線，無線を問わない各種の公衆回線網，専用線網や、ＬＡＮ等の各種のネットワークが含まれる。

【００９５】

データベース９３は、ネットワーク９２に接続されており、回答者コンピュータ

９１１～９１Ｎからネットワーク９２を通じて送信された、複数のアンケート回答者からのアンケート回答文を蓄積する。

【００９６】

テキスト分類エンジン９４は、ルール出力手段１９が分析結果のルール形式の知識をネットワーク９２を通じて送信できる点を除いては、図１に示した第１の実施の形態に係るアンケート回答分析システムにおけるテキスト分類エンジン１４と全く同様のものである。したがって、対応する部分には同一符号を付して、それらの詳しい説明を省略する。

【００９７】

要求者コンピュータ９５は、ネットワーク９２を通じてテキスト分類エンジン９４に分析結果のルール形式の知識を要求し、テキスト分類エンジン９４から分析結果のルール形式の知識をネットワーク９２を通じて受信することができる。

【００９８】

次に、このように構成された第３の実施の形態に係るアンケート回答分析システムの動作について説明する。

【００９９】

アンケート回答者は、回答者コンピュータ９１１～９１Ｎからネットワーク９２を介して自然言語による自由回答記述を含むアンケート回答文を送る。なお、アンケート回答者は、Ｎ人とする。

【０１００】

データベース９３は、ネットワーク９２に接続されており、複数のアンケート回答者からのアンケート回答文を蓄積する。

【０１０１】

テキスト分類エンジン９４は、データベース９３から複数のアンケート回答文を読み出して、アンケート回答文を分類するルール形式の知識を導き出す。テキスト分類エンジン９４は、ネットワーク９２に接続し、分析結果のルール形式の知識を要求者コンピュータ９５の要求に応じてネットワーク９２を通じて配信する。なお、テキスト分類エンジン９４の動作の詳細は、分析結果のルール形式の知識がネットワーク９２を通じて配信される点を除いて、第１の実施の形態に係る

アンケート回答分析システムにおけるテキスト分類エンジン１４の場合と全く同様であるので、その詳しい説明を割愛する。

【０１０２】

第３の実施の形態にアンケート回答分析システムは、例えば、以下のようなビジネスとして展開できる。企業に関するイメージ調査，特定の商品やサービス等のアンケートを請け負って、図２のようなアンケート項目でアンケートをネットワーク９２上で実施し、ネットワーク９２を通じてオンラインで集められた自然言語による自由回答記述を含むアンケート回答文をデータベース９３で蓄え、そこからアンケート回答文を呼び出して、テキスト分類エンジン９４を使用することにより得られる分析結果を、要求があれば、ネットワーク９２を通じて要求者に配信サービスすることによりビジネスを行う。

【０１０３】

（４）第４の実施の形態

図１０は、本発明の第４の実施の形態に係るアンケート回答分析システムの全体構成を示すブロック図である。本実施の形態に係るアンケート回答分析システムは、図１に示した第１の実施の形態に係るアンケート回答分析システムにおいて、データベース１３に接続されたコンピュータ１０１にテキスト分類エンジンプログラムを記録した記録媒体１０２を備えるようにしたものである。その他の構成は、第１の実施の形態に係るアンケート回答分析システムと全く同様であるので、対応する部分には同一符号を付してそれらの詳しい説明を省略する。

【０１０４】

このように構成された第４の実施の形態に係るアンケート回答分析システムでは、記録媒体１０２からコンピュータ１０１にテキスト分類エンジンプログラムが読み込まれ、形態素解析手段１５，カテゴリ・テキスト指定手段１６，属性選択手段１７，ルール学習手段１８，およびルール出力手段１９を含むテキスト分類エンジン１４としてコンピュータ１０１の動作を制御する。コンピュータ１０１上でのテキスト分類エンジン１４の動作の詳細は、第１の実施の形態に係るアンケート回答分析システムの場合と全く同様になるので、その詳しい説明を割愛する。

【０１０５】

（５）第５の実施の形態

図１１は、本発明の第５の実施の形態に係るアンケート回答分析システムの全体構成を示すブロック図である。本実施の形態に係るアンケート回答分析システムは、図８に示した第２の実施の形態に係るアンケート回答分析システムにおいて、データベース８２に接続されたコンピュータ１１１にテキスト分類エンジンプログラムを記録した記録媒体１１２を備えるようにしたものである。その他の構成は、第２の実施の形態に係るアンケート回答分析システムと全く同様であるので、対応する部分には同一符号を付してそれらの詳しい説明を省略する。

【０１０６】

このように構成された第５の実施の形態に係るアンケート回答分析システムでは、記録媒体１１２からコンピュータ１１１にテキスト分類エンジンプログラムが読み込まれ、形態素解析手段１５，カテゴリ・テキスト指定手段１６，属性選択手段１７，ルール学習手段１８，およびルール出力手段１９を含むテキスト分類エンジン８３としてコンピュータ１１１の動作を制御する。コンピュータ１１１上でのテキスト分類エンジン８３の動作の詳細は、第２の実施の形態に係るアンケート回答分析システムの場合と全く同様になるので、その詳しい説明を割愛する。

【０１０７】

（６）第６の実施の形態

図１２は、本発明の第６の実施の形態に係るアンケート回答分析システムの全体構成を示すブロック図である。本実施の形態に係るアンケート回答分析システムは、図９に示した第３の実施の形態に係るアンケート回答分析システムにおいて、データベース９３に接続されたコンピュータ１２１にテキスト分類エンジンプログラムを記録した記録媒体１２２を備えるようにしたものである。その他の構成は、第３の実施の形態に係るアンケート回答分析システムと全く同様であるので、対応する部分には同一符号を付してそれらの詳しい説明を省略する。

【０１０８】

このように構成された第６の実施の形態に係るアンケート回答分析システムでは

、記録媒体１２２からコンピュータ１２１にテキスト分類エンジンプログラムが読み込まれ、形態素解析手段１５，カテゴリ・テキスト指定手段１６，属性選択手段１７，ルール学習手段１８，およびルール出力手段１９を含むテキスト分類エンジン９４としてコンピュータ１２１の動作を制御する。コンピュータ１２１上でのテキスト分類エンジン９４の動作の詳細は、第３の実施の形態に係るアンケート回答分析システムの場合と全く同様になるので、その詳しい説明を割愛する。

【０１０９】

【発明の効果】

【０１１０】

本発明の第１の効果は、企業に関するイメージ調査，特定の商品やサービス等のアンケートを請け負って、アンケートをネットワーク上で実施し、ネットワークを通じてオンラインで集められた自然言語による自由回答記述を含むアンケート回答文をデータベースに蓄え、そこからアンケート回答文を呼び出して、テキスト分類エンジンを使用することにより得られるルール形式の知識を分析結果として要求者に販売できることである。

【０１１１】

本発明の第２の効果は、企業に関するイメージ調査，特定の商品やサービス等のアンケートを請け負って、アンケートを実施し、自然言語による自由回答記述を含むアンケート回答文を一度に回収して、これをデータベースで蓄え、そこからアンケート回答文を呼び出して、テキスト分類エンジンを使用することにより得られるルール形式の知識を分析結果として要求者に販売できることである。

【０１１２】

本発明の第３の効果は、企業に関するイメージ調査，特定の商品やサービス等のアンケートを請け負って、アンケートをネットワーク上で実施し、ネットワークを通じてオンラインで集められた自然言語による自由回答記述を含むアンケート回答文をデータベースで蓄え、そこからアンケート回答文を呼び出して、テキスト分類エンジンを使用することにより得られるルール形式の知識を分析結果として、要求者からの要求に応じてネットワークを通じて配信サービスすることによ

り販売できることである。

【図面の簡単な説明】

【図１】

本発明の第１の実施の形態に係るアンケート回答分析システムの構成を示すブロック図である。

【図２】

図１中のデータベースに蓄積されたアンケート回答文を例示する図である。

【図３】

図１中のテキスト分類エンジンにおける処理を示すフローチャートである。

【図４】

図３中の属性選択ステップのより詳細な処理を示すフローチャートである。

【図５】

図３中のルール学習ステップのより詳細な処理を示すフローチャートである。

【図６】

図１中のテキスト分類エンジンによる分析結果であるルール形式の知識（確率的決定リスト）の一例を示す図である。

【図７】

図１中のテキスト分類エンジンによる分析結果であるルール形式の知識（確率的決定リスト）の他の例を示す図である。

【図８】

本発明の第２の実施の形態に係るアンケート回答分析システムの構成を示すブロック図である。

【図９】

本発明の第３の実施の形態に係るアンケート回答分析システムの構成を示すブロック図である。

【図１０】

本発明の第４の実施の形態に係るアンケート回答分析システムの構成を示すブロック図である。

【図１１】

本発明の第５の実施の形態に係るアンケート回答分析システムの構成を示すブロック図である。

【図１２】

本発明の第６の実施の形態に係るアンケート回答分析システムの構成を示すブロック図である。

【符号の説明】

１２　ネットワーク

１３　データベース

１４　テキスト分類エンジン

１５　形態素解析手段

１６　テキスト・カテゴリ指定手段

１７　属性選択手段

１８　ルール学習手段

１９　ルール出力手段

３１　形態素解析ステップ

３２　テキストおよびカテゴリ指定ステップ

３３　属性選択ステップ

３４　ルール学習ステップ

３５　ルール出力ステップ

４１　ΔＳＣ（ω）計算ステップ

４２　属性選択ステップ

５１　データ成形ステップ

５２　成長処理ステップ

５３　刈り込み処理ステップ

８１　アンケート回答入力手段

８２　データベース

８３　テキスト分類エンジン

９２　ネットワーク

９３　データベース

９４　テキスト分類エンジン

９５　要求者コンピュータ

１０１，１１１，１２１　コンピュータ

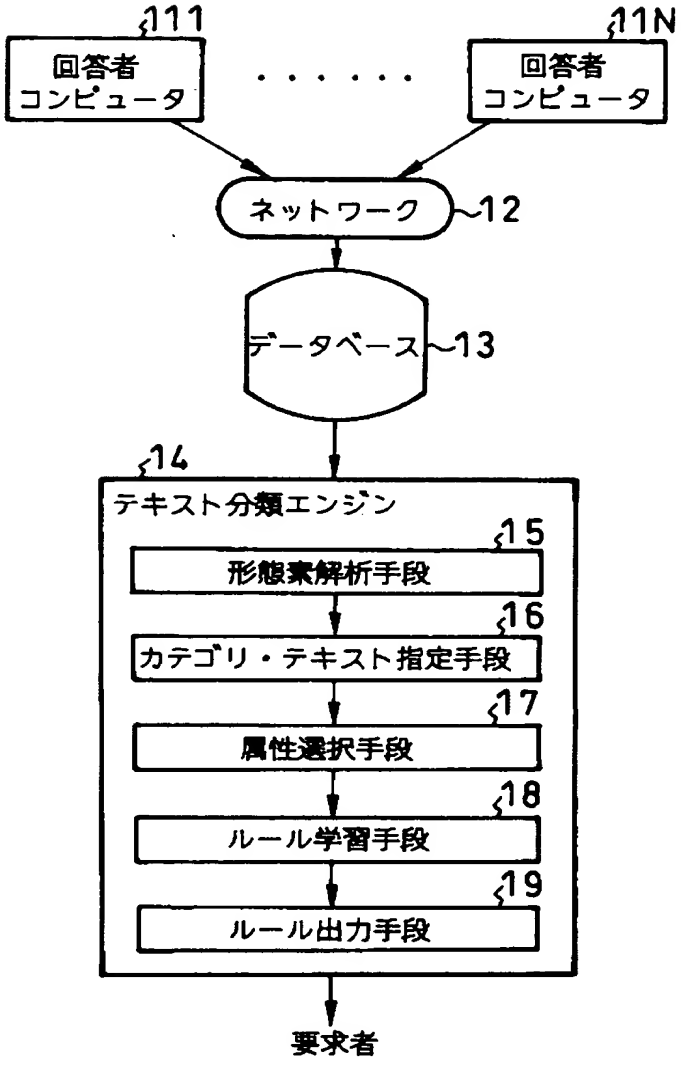
１０２，１１２，１２２　記録媒体

１１１～１１Ｎ　回答者コンピュータ

９１１～９１Ｎ　回答者コンピュータ

【書類名】　図面

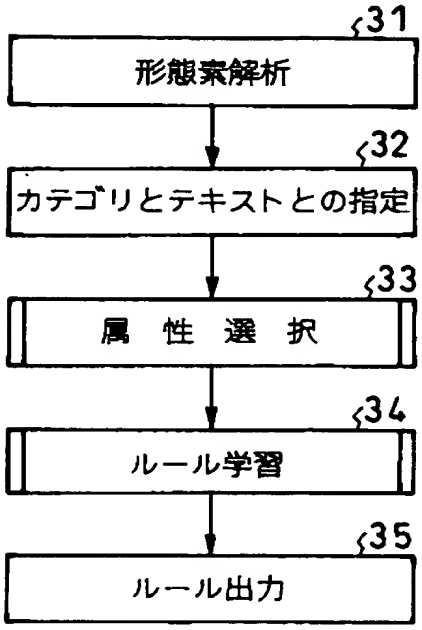
【図１】



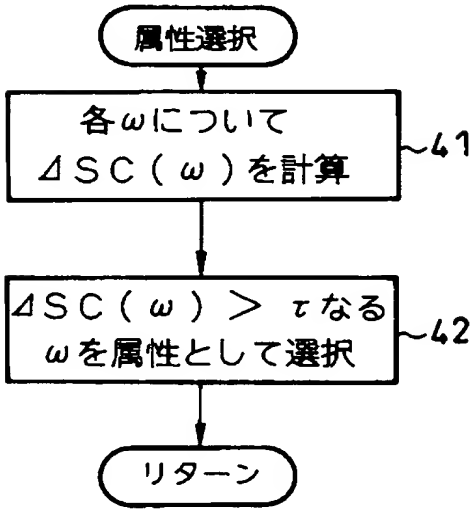

【図２】

| アンケート回答者 | ハイテク企業としてどこを想起しますか？ | あなたにとってハイテクとは何ですか？ | ハイテク商品として何を思い浮かべますか？ | ・・・・・ |
|---|---|---|---|---|
| 1 | A社 | 進化した未来的な機械 | ロボット | ・・・・・ |
| 2 | C社 | 使いやすく優しいもの | 携帯電話 | ・・・・・ |
| 3 | A社 | 高速で高性能な機械 | パソコン | ・・・・・ |
| ・・・ | ・・・ | ・・・・・ | ・・・・・ | ・・・・・ |

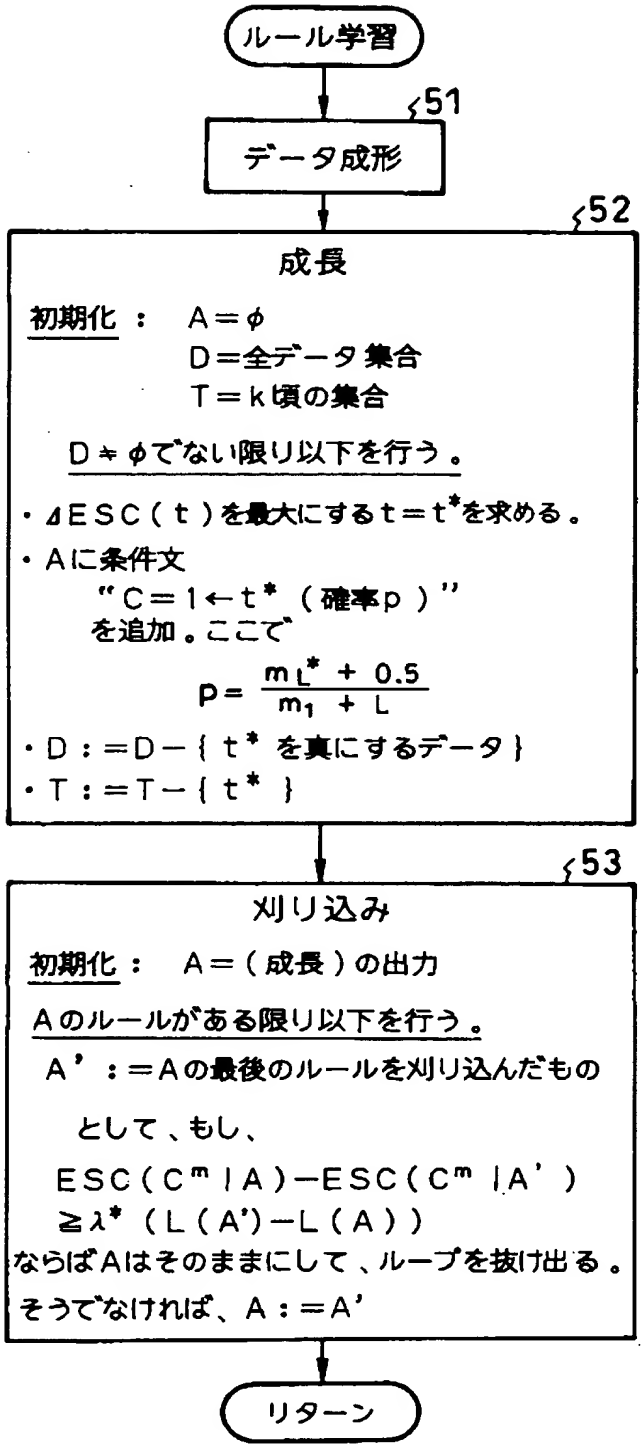
【図３】


31
形態素解析
32
カテゴリとテキストとの指定
33
属　性　選　択
34
ルール学習
35
ルール出力

【図４】


属性選択
各ωについて
ΔSC（ω）を計算
41
ΔSC（ω）＞ τなる
ωを属性として選択
42
リターン

【図５】


ルール学習
51
データ成形
52
成長
初期化 ： A＝φ
D＝全データ集合
T＝k項の集合
D≠φでない限り以下を行う。
・ΔESC(t)を最大にするt＝t*を求める。
・Aに条件文
"C＝1←t*（確率p）"
を追加。ここで
$p=\frac{m_L^* + 0.5}{m_1 + L}$
・D：＝D－{t*を真にするデータ}
・T：＝T－{t*}
53
刈り込み
初期化 ： A＝（成長）の出力
Aのルールがある限り以下を行う。
A'：＝Aの最後のルールを刈り込んだもの
として、もし、
$ESC(C^m|A)-ESC(C^m|A') \geq \lambda^*(L(A')-L(A))$
ならばAはそのままにして、ループを抜け出る。
そうでなければ、A：＝A'
リターン

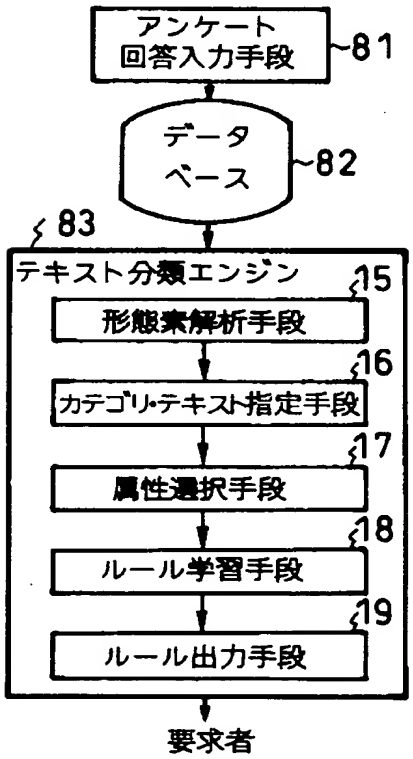
【図６】

A社 ← 使いやすい〔92.0%〕
A社 ← 未来&プライベート〔87.2%〕
A社 ← 疲労&軽減〔78.0%〕
A社 ← 簡単だ〔65.8%〕
A社 ← 楽しい〔56.2%〕
A社でない ← それ以外〔79.4%〕

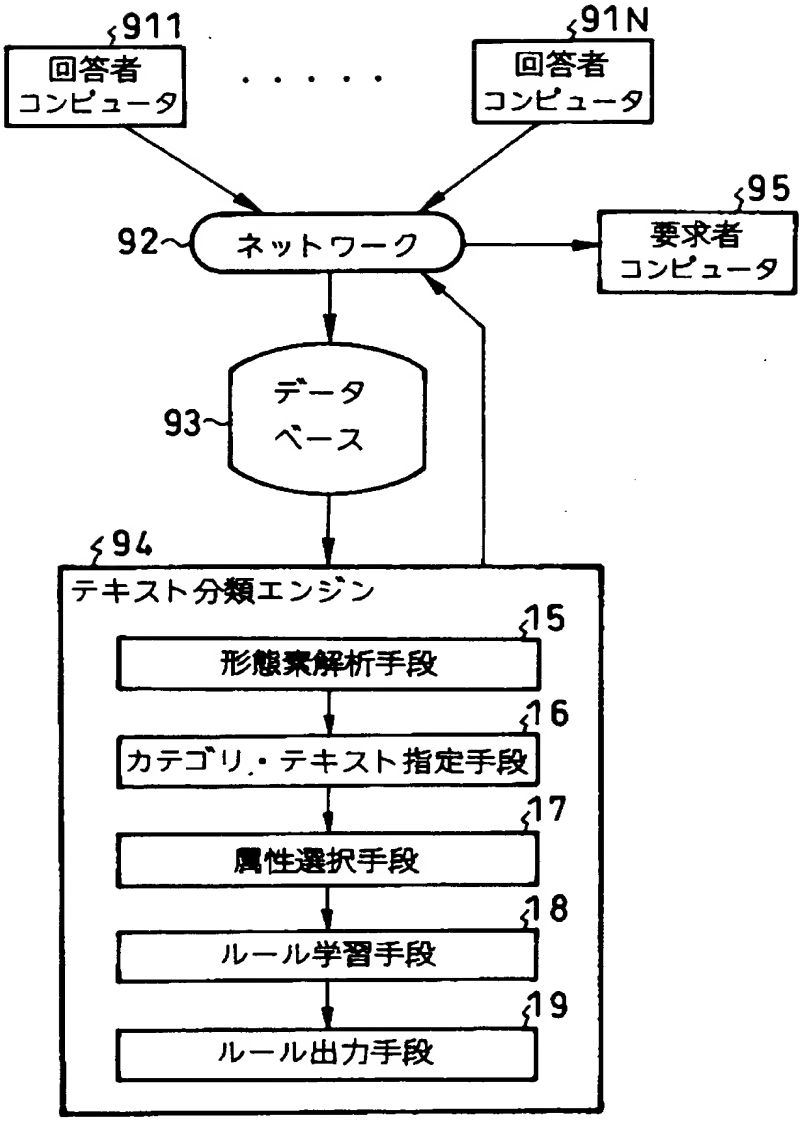
【図７】

B社 ← 迅速だ〔82.0%〕
B社 ← 機械&効率〔77.8%〕
B社 ← 機械&操作〔76.0%〕
B社 ← 巧妙だ〔60.8%〕
B社 ← 優れた〔60.2%〕
B社でない ← それ以外〔76.4%〕

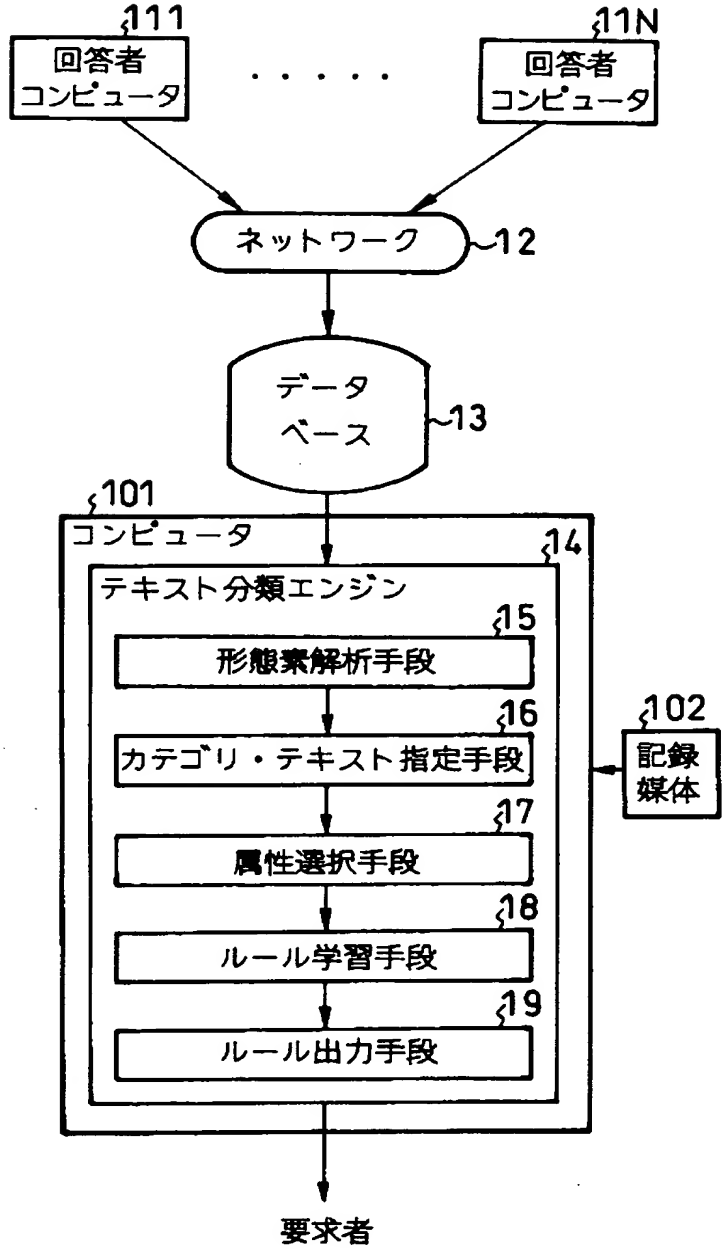
【図８】



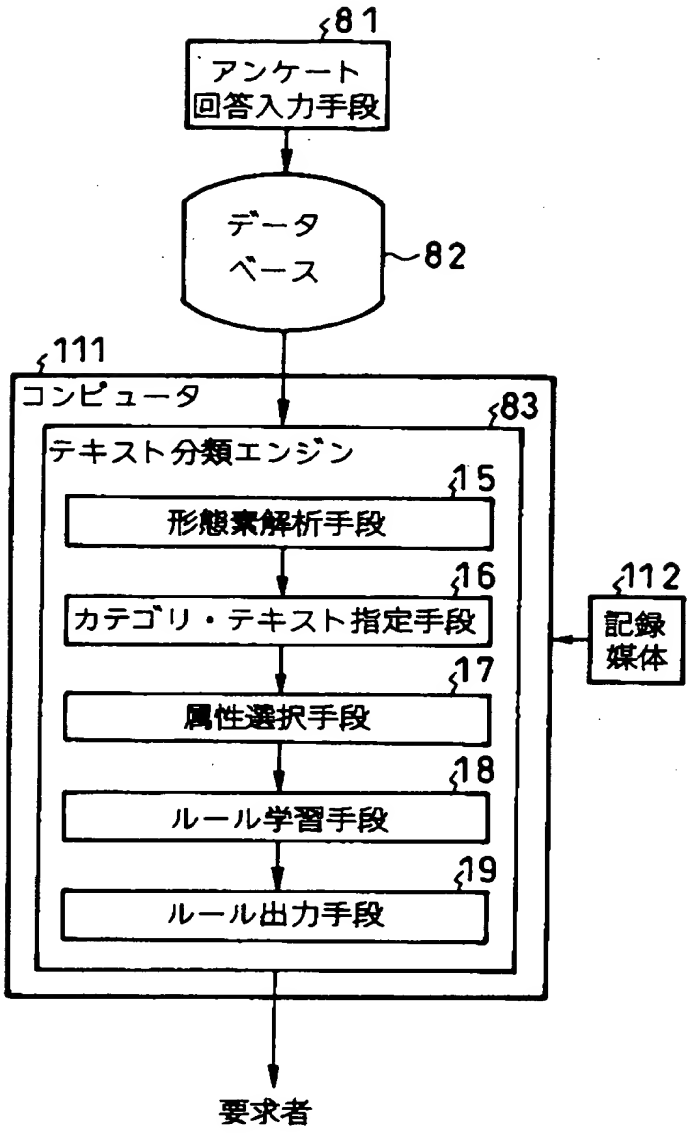
【図９】


911
回答者
コンピュータ
91N
回答者
コンピュータ
95
92
ネットワーク
要求者
コンピュータ
93
データ
ベース
94
テキスト分類エンジン
15
形態素解析手段
16
カテゴリ・テキスト指定手段
17
属性選択手段
18
ルール学習手段
19
ルール出力手段

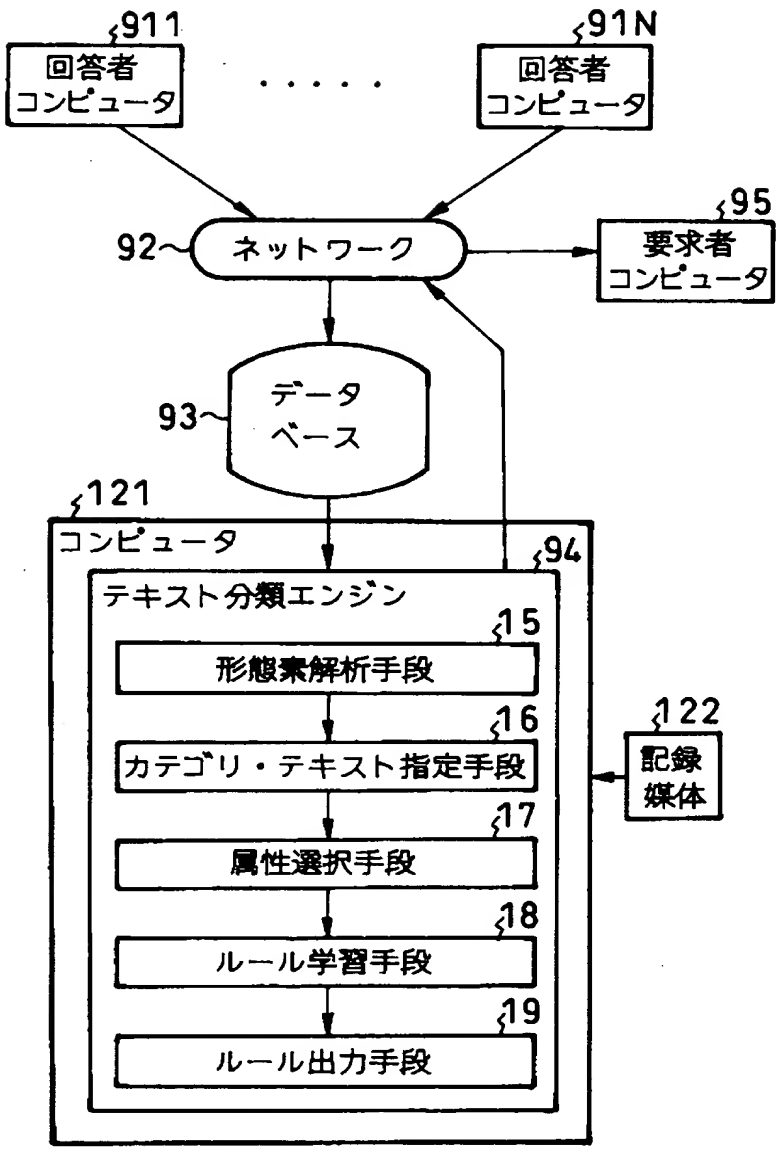
【図１０】


111
回答者
コンピュータ
・・・・・
11N
回答者
コンピュータ
ネットワーク
12
データ
ベース
13
101
コンピュータ
14
テキスト分類エンジン
15
形態素解析手段
16
カテゴリ・テキスト指定手段
17
属性選択手段
18
ルール学習手段
19
ルール出力手段
102
記録
媒体
要求者

【図１１】


81
アンケート
回答入力手段
データ
ベース
82
111
コンピュータ
83
テキスト分類エンジン
15
形態素解析手段
16
カテゴリ・テキスト指定手段
17
属性選択手段
18
ルール学習手段
19
ルール出力手段
112
記録
媒体
要求者

【図１２】


911
回答者
コンピュータ
91N
回答者
コンピュータ
95
92
ネットワーク
要求者
コンピュータ
93
データ
ベース
121
コンピュータ
94
テキスト分類エンジン
15
形態素解析手段
16
カテゴリ・テキスト指定手段
17
属性選択手段
18
ルール学習手段
19
ルール出力手段
122
記録
媒体

【書類名】　要約書

【要約】

【課題】ネットワークを通じて回収した自然言語による自由回答記述を含むアンケート回答文から、テキスト分類エンジンを用いることにより、アンケート回答分析を自動的に行い、分析結果をルール形式の知識として要求者に出力する。

【解決手段】アンケート回答者は回答者コンピュータ１１１～１１Ｎからアンケート回答文を送信する。アンケート回答文は、ネットワーク１２を通じてデータベース１３に蓄積される。テキスト分類エンジン１４は、データベース１３から蓄積されたアンケート回答文を読み出して、アンケート回答文を分類するルールを学習して、要求者に出力する。

【選択図】　図１

# 認定・付加情報

| | |
|---|---|
| 特許出願の番号 | 特願２０００－０７１６５７ |
| 受付番号 | ５００００３０７６５９ |
| 書類名 | 特許願 |
| 担当官 | 第七担当上席　００９６ |
| 作成日 | 平成１２年　３月１６日 |

＜認定情報・付加情報＞

【提出日】　平成12年 3月15日

次頁無

# 出 願 人 履 歴 情 報

識別番号　〔０００００４２３７〕

１．変更年月日　１９９０年　８月２９日
［変更理由］　新規登録
住　所　東京都港区芝五丁目７番１号
氏　名　日本電気株式会社